Response to SC1:

We thank Dr. León-Luis et al. for their very helpful comments and fast response to the issue that we identified and reported in our paper. The results provided in the document are very important and we have adapted some of this information into the revised paper. Our responses are given below in black with the comments from León-Luis et al. in blue. The new/revised text in the modified manuscript is given in red (italicized).

1 Comparison with the RBBC-E triad in León-Luis *et al.* (2018)

In the paper the authors claim that the comparison with the RBCC-E Triad presented in León-Luis *et al.* (2018) should not be carried out because the calculation is not consistent with the results of Model 1 in the present paper by Zhao et al. Note Model 1 was proposed in Fioletov *et al.* (2015).

In León-Luis *et al.* (2018), we calculate a quadratic polynomial fit for every Brewer as

$$O_3 = A + B \cdot (t - t_0) + C \cdot (t - t_0)^2 \quad (1)$$

obtaining for each instrument the corresponding values of A, B and C. Model 1 in Fioletov *et al.* (2015) however calculates common B and C values for all instruments.

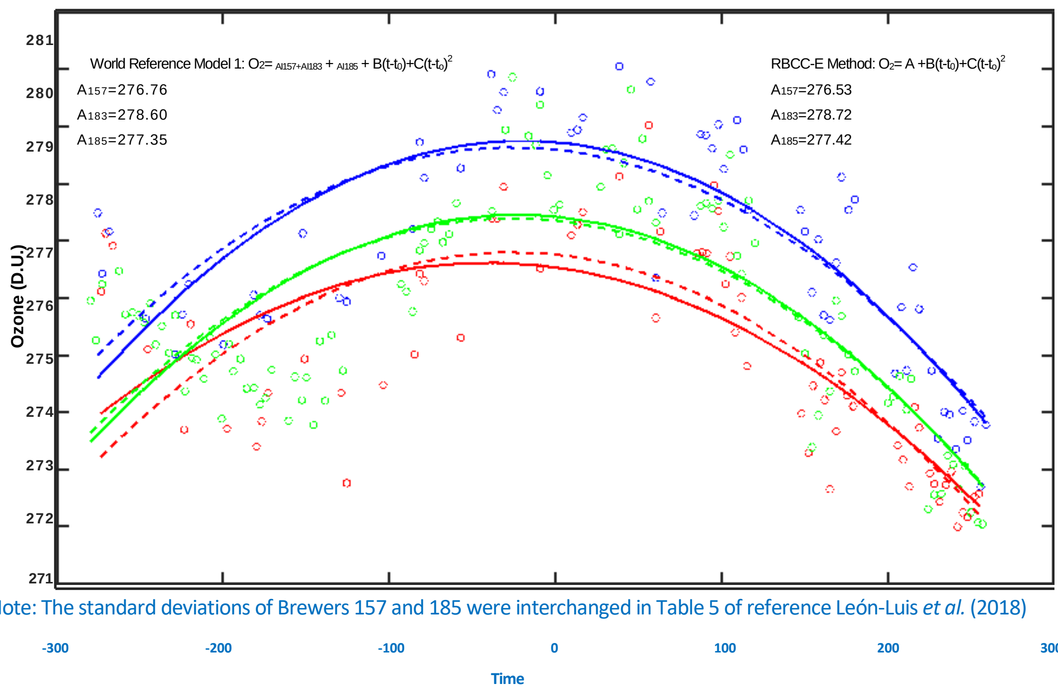
We take the opportunity of the open discussion of this paper to update the calculations of the RBCC-E Triad to be consistent with Fioletov *et al.* (2015), and also to compare the results of both Eq. 1 and Model 1 from Fioletov *et al.* (2015).

Table 1 contains the 3-month standard deviation of the A_i coefficients obtained when the RBCC-E data are re-evaluated using Model 1, together with our previous published results. As can be observed, the values for each Brewer change slightly, depending on the method applied. However, the mean value of the Triad is similar, 0.23% versus 0.27%. This result confirms that there is very little difference between both methods when are applied to the RBCC-E Triad data.

This point can be better understood with an example. Fig. 1 demonstrates the total ozone column recorded on November 16th, 2016 (Fig. 4 in León-Luis *et al.* (2018)), where the data have been fitted used the two methods previously described. Table 2 contains the A, B and C coefficients calculated by both methods. As can be seen, regardless of the method used,

Table 1. RBCC-E and World Reference Triads: 3-month standard deviation. We include the values of the World Reference Triad from Zhao *et al.* (2020) for comparison.

RBCC-E			World Reference			
Brewer	σ_{3month}, [Eq.1]	σ_{3month}, Model 1	Brewer	σ_{3month}, Model 1	Brewer	σ_{3month}, Model 1
#157	0.20 [1]	0.19	#008	0.43 (0.40)	#145	0.44
#183	0.31	0.26	#014	0.36 (0.46)	#187	0.26
#185	0.29	0.23	#015	0.42 (0.39)	#191	0.33
Mean	0.27	0.23		0.40 (0.42)		0.34



Note: The standard deviations of Brewers 157 and 185 were interchanged in Table 5 of reference León-Luis *et al.* (2018)

Figure 1. Ozone values measured on November 16th, 2016, marked with circles. Solid and dotted lines correspond to the 2nd order polynomial fitted using Eq. 1 (RBCC-E method) and Model 1 from Fioletov *et al.* (2015) (World Reference Model method), the Time units are the minutes from solar noon. The *A* coefficients calculated with both methods are also shown.

the derived *A* coefficients are very similar. Therefore, the mean daily value of the RBCC-E Triad, the relative errors for each instrument, and the standard deviation, calculated from these coefficients, should not differ significantly. Furthermore, the *B* and *C* coefficients calculated by both methods are similar, which suggests that the adjusted functions will exhibit the same behavior as shown the Fig. 1. In conclusion, although both calculation methods are not the same, the results in the case of the RBCC-E Triad are very close. A similar result is achieved when no mathematical adjustment is used and the mean from the simultaneous measurements is calculated directly.

2

Table 2. Coefficients calculated with the two methods for the RBCC-E Triad data of November 16th, 2016 A, B, and C coefficients

RBCC-E	$A_{157} = 276.53$, $B_{157} = -0.0040$, $C_{157} = -4.855e-5$ $A_{183} = 278.72$, $B_{183} = -0.0025$, $C_{183} = -6.337e-5$ $A_{185} = 277.42$, $B_{185} = -0.0028$, $C_{185} = -6.033e-5$
World Reference	$A_{157} = 276.76$, $A_{183} = 278.60$, $A_{185} = 277.35$ $B = -0.0030$, $C = -5.8122e-5$

Table 3. Percentage difference of the mean of the three instruments, mean and its standard deviation and the percentage of observations 1% 0.5% and 0.25% of the five minutes simultaneous measurements and daily mean

	Brewer	Mean	σ	<1%	<0.5%	<0.25%
5 min	#157	-0.041	0.342	0.994	0.909	0.687
	#183	0.023	0.372	0.991	0.900	0.701
	#185	0.018	0.342	0.99	0.921	0.758
daily	#157	-0.002	0.245	0.999	0.979	0.816
	#183	-0.005	0.309	0.999	0.931	0.757
	#185	0.007	0.267	0.992	0.954	0.866

Table 3 shows the mean and standard deviation of the ratios for the 5 minutes simultaneous measurements and daily mean values and note that the standard deviations values in this table are fairly similar to those in Table 1, even though the periods used for the calculations are not the same (5 minutes, daily and 3 months).

Up to this point, we have shown that both methods produce a similar result. The difference between the standard deviation reported by both Brewer Triads could then be associated to others factors which have not been considered in these works, such as e.g. the intra-day ozone variability or the number of ozone (Direct Sun) measurements made per day at each station.

These factors can affect the robustness of the mathematical fits and, hence, introduce small differences between the calculated *A* coefficients that are difficult to evaluate for two stations so far apart.

We thanks the Izana Atmospheric Research Centre group from AEMET for providing such detailed recalculations for the RBCC-E data, with the accurate method proposed in Fioletov et al., 2005. We fully agree with the group that using the index matrix method or the simple second-order fitting would not alter the results of the estimated random errors for Brewer triads. As the most accurate and precise total ozone observation instruments, well-maintained Brewers can have less than 0.5% random uncertainties. The difference caused by the selected fitting algorithm should be well within this uncertainty levels for most of the cases. We pointed this issue out due to two reasons, 1) a part of the results (in León-Luis et al., 2018) was referred to as the use of Model 1 designed in Fioletov et al., 2005, which is not accurate. We thank the AEMET group for providing these re-calculated comparable results. 2) Moreover, the design of Model 1 is just a part of the whole evaluation scheme that been proposed. Model 2 designed in Fioletov et al. 2005 needs a "baseline" ozone, which is calculated from Model 1. For cases illustrated by this comments/report (i.e., RBCC-E data in 2016 Nov. 16), the baseline for this day can be defined as $(A_{157} + A_{183} + A_{185})/3 + B(t-t_0) + C(t-t_0)^2$, if B and C term of the fitting is "common" factors shared by all three instruments. However, when one selects to use a simple second-order fit for each of the instrument, then one will have three B terms and three C terms. One may argue that we also can average B and C terms to receive a "baseline" ozone; however, for some case, the B and C terms can be very different from instruments to instruments (if one only apply the simple fitting). We want to emphasize that the design of Model 1 is only a starting part of the evaluation scheme. The designed models work together to provide a guide in evaluating the performance of Brewers, if no other high-quality reference data can be used as a "referee".

2 Additional comments

In this section we include some other comments to Zhao et al. (2020), but first we want to acknowledge the effort of the World Reference Triad to maintain all these instruments during decades with such a high precision. Once the precision of the Triads has been established the challenge is to quantify the uncertainty, especially that produced by the described instrumental issues and include them in the analysis.

We thank the AEMET group again for providing us with such important comments and suggestions. The collaborations within the Brewer network is important not just for Brewer researchers, but also for the global ozone monitoring activities and related ozone research studies. The ECCC group designed and maintained the Brewer instruments since the 1970s. Almost a half-century of dedications to ozone monitoring work is a big accomplishment made by all Brewer scientists, technicians, and managers, and more importantly from our collaborators. As noted in the table of world reference instruments' primary calibration trips, some of the reference instruments were calibrated at Izana with great help from the AEMET group.

1. We do not agree with the comment that the 0.5% level cannot be achieved due to limitations of the Brewer hardware. Some of the issues described, such as for example the filter non linearity, can be addressed, and indeed are accounted for

in the processing performed at Eubrewnet. Eubrewnet's processing also takes into account the issue described for Brewer #15 – the observations not compensated with mercury tests are automatically filtered out.

The purpose of this work is to evaluate the triad performance based on the existing calibration results. Numerous works have been done by researchers worldwide in past decades to improve Brewers' accuracy and precision. However, not all of them have been implemented to the current reference instruments' results. We also provided a more detailed reply on issues such as filter non-linearity correction to referee #2. ECCC group maintains the largest number of Brewers within this community (i.e. more than 40 Brewers). Meet the WMO/GAW requirement and performing centralized data processing with minimal intervention are critical to the ECCC Brewer programme. In summary, the goal of this study is to evaluate the overall performance of the current long historical triad record, and any further data improvement can be performed when a higher-precision reprocessing is needed and called upon by WMO/GAW. ECCC group welcomes and is looking forward to continue and further our collaborations with AEMET group on these activities in future.

2. The cited Pandora manual has more than 150 pages, so it is difficult to find the ozone processing details. It could be better to refer to the ozone processing in the user guidelines avaliable at https://www.pandonia-global-network.org/wp-content/uploads/2020/01/LuftBlick_FRM4AQ_PGNUserGuidelines_RP_2019009_v1.pdf. Furthermore, if we understand it correctly, the data used in the present paper by Zhao et al. is not the operational one that is available to the public for download.

We have included this in the reference. The effective temperature corrected Pandora ozone is not available on the PGN website. We have upload the corrected TCO data to ECCC's public data server and can be downloaded from: https://collaboration.cmc.ec.gc.ca/cmc/arqi/Zhao_et_al_amt-2020-324/

Additional information on Pandora calibrations, operation, retrieval algorithms and correction method can be found in Cede (2019; Cede et al., 2019), Tzortziou et al., (2012), and Zhao et al., (2016).

Cede, A., Tiefengraber, M., Gebetsberger, M. and Kreuter, M.: TN on PGN products "correct use" guidelines, Pandonia Global Network. [online] Available from: https://www.pandonia-global-network.org/wp-content/uploads/2020/01/LuftBlick_FRM4AQ_PGNUserGuidelines_RP_2019009_v1.pdf (Accessed 13 November 2020), 2019.

3. It looks that there is a trend on the Merra comparison from 2005 to 2015, with Brewer #015 going from +2% to -2%

We are not sure Brewer #015 is the only instrument that shows a “trend”. For example, Brewers #008 and #014 look like they also have a decreasing trend in this period, just not as “strong” as the one observed by Brewer #015. It is difficult for us to determine if this trend is real. To continue this investigation, one will need to collaborate with the reanalysis group and investigate not just Brewers, but also any upgrade or modifications in the model and its source of assimilation. However, we think that this interesting topic could be a standalone research topic and is beyond the scope of the current study.

4. Appendix A. The standard deviation of the ozone measurement is strongly affected by clouds and is also used as cloud mask to filter the AOD measurements affected by rapid moving clouds (López-Solano et al., 2017). Some of the brewer are equipped with full sky cameras, are the observations reported in Zhao et al. (2020) also filtered by clouds?

The standard data screening procedure used the standard deviations of 5 individual Brewer measurements to remove DS values obtained under cloudy conditions or under moving clouds. In very rare occasions when the standard deviations of such measurements are within the established limits, but the Sun is obscured by the clouds, an additional screening is done using the absolute intensity at the longest wavelength (320nm). Sky cameras are not very useful for such screening since the sky cloud coverage does not really affects DS measurements unless the clouds block the Sun. Sky camera data was not used in this study.

5. Appendix A. Figure A2 shows the dependence with the ozone air mass factor (AMF), as the stray-light is a function of AMF (Karppinen et al. (2015)) , but in the text the discussion is focused on the solar zenith angle and air mass

Further discussions of the stray light issue were provided in our answers to referees #1 and #2. Please refer to our answers in those replies.

6. Appendix A. An statistical approach to estimate the single triad stray light Diemoz et al. (2015) or the determination of the empirical correction by comparison with the double one (Redondas et al., 2018) could be performed to the dataset.

We thank the AEMET group for this very useful suggestion. When re-processing of the Brewer reference instruments records are needed, we will try to implement these proposed methods.